'I did it for the LULZ': How the dark personality predicts online disinhibition and aggressive online behavior in adolescence

Author

Kurek, Anna, Jose, Paul E, Stuart, Jaimee

Published

2019

Journal Title

Computers in Human Behavior

Version

post-print

DOI

10.1016/j.chb.2019.03.027

Accepted Manuscript

‘I did it for the LULZ’: How the dark personality predicts online disinhibition and aggressive online behavior in adolescence

Anna Kurek, Paul E. Jose, Jaimee Stuart

PII:	S0747-5632(19)30125-6
DOI:	10.1016/j.chb.2019.03.027
Reference:	CHB 5963
To appear in:	*Computers in Human Behavior*
Received Date:	06 January 2018
Accepted Date:	23 March 2019

Please cite this article as: Anna Kurek, Paul E. Jose, Jaimee Stuart, ‘I did it for the LULZ’: How the dark personality predicts online disinhibition and aggressive online behavior in adolescence, *Computers in Human Behavior* (2019), doi: 10.1016/j.chb.2019.03.027

'I did it for the LULZ': How the dark personality predicts online disinhibition and aggressive online behavior in adolescence

[1]Anna Kurek,

[1]Paul E. Jose,

and [2]Jaimee Stuart

[1]Victoria University of Wellington

[2]Griffith University

Author note

Corresponding author: Anna Kurek, P.O. Box 600, Victoria University of Wellington, Wellington, New Zealand 6012; anna.kurek@vuw.ac.nz; fax: 0064-04-463-6023.

Appreciation is expressed to the schools, teachers, and students who participated in the study, and to Victoria University of Wellington for helping to fund this study.

Abstract

A large proportion of youth believe that the world of cyberspace provides them with a relatively safe and anonymous digital bubble ripe for uninhibited self-expression. At the same time, observers have noted an increase of individuals behaving in an unrestrained manner on the Internet, while researchers have reported elevated rates of cyber aggressive behavior. What remains unclear, however, is whether, and how, disinhibition might be related to cyber aggression. In an aim to explore the possible associations, a large sample (total N = 709) of high school (M_{age} = 15.56 years) respondents from New Zealand were recruited, and completed a survey featuring scales assessing personality and technology behaviors, attitudes, habits, and trends. The present study was designed to investigate whether the three dark personality traits of narcissism, psychopathy, and sadism would predict false self perceptions, and in sequence, online disinhibition and aggressive online behavior. All three dark personality traits, as well as false self, were positively associated with online disinhibition. Perceptions of false self were found to be a significant predictor of cyber aggression when mediated by online disinhibition. In the case of cyber aggression, however, psychopathy, sadistic traits, and online disinhibition were found to be significant predictors of this outcome. The results collectively provide a more nuanced understanding of how antisocial personality traits are associated with maladaptive identity formation (i.e., endorsement of false self beliefs) as well as maladaptive online behavior.

Keywords: *Narcissism; Sadism; Psychopathy; Identity; Online Disinhibition; Cyber Aggression*

'I did it for the LULZ': How the dark personality predicts online disinhibition and aggressive online behavior in adolescence

Understanding the Roots of Cyber Aggression

As the online social world has expanded, some observers have noted an increase of individuals behaving in an unrestrained manner on the Internet. Identified as a considerable problem, particularly among adolescents (Kowalski, Giumetti, Schroeder, & Lattanner, 2014), this growing phenomenon seemingly coincides with the drastic increase in the prevalence of aggressive online behavior (Garett, Lord, & Young, 2016; Rawhide, 2017). Eager to unearth the complex motivations behind the burgeoning culture of cyber hostility, researchers have argued that the unique and continuously changeable digital landscape now demands a broader understanding of aggressive online behavior that goes beyond the narrow scope of 'online bullying' (Corcoran, McGuckin, & Prentice, 2015).

Much of the current speculation behind the increase of aggressive digital behavior in youth coincides with the idea that the Internet is an important anonymous space that promotes honest self-expression free of immediate judgement or consequence (Suler, 2004). Indeed, in many ways, this digital landscape provides a context in which there is an absence of important social cues concerning appropriate behavior (which are generally present in face-to-face interactions), that allow youth to engage in hostile or antisocial behavior online (Hemphill & Heerde, 2014; Pornari & Wood, 2010). Because victims' responses may be absent, suppressed, or delayed on the Internet, it is possible that aggressive individuals fail to accurately predict the severity of harm caused by their actions.

Moreover, increased opportunities for aggression and the ability or willingness to override inhibition are other common factors that have been used to explain why certain individuals are more likely to act aggressively online (Anderson & Bushman, 2002). Some research has suggested that increases in aggressive behavior can be linked to biases in morality, specifically distorted moral reasoning, that influences the individual to minimize feelings of guilt or remorse (Hymel, & Bonanno, 2014; Malti, Gasser, & Gutzwiller-Helfenfinger, 2010; Menesini, Nocentini, & Camodeca, 2013; Pozzoli, Gini, & Thornberg, 2016). Whereas other studies have suggested that fluctuations in self-image, self-concept, self-standards, and sense of self-worth are related to the increased likelihood of aggressive acts online (Crocker & Wolfe, 2001; Ronningstam, 2017). Despite these promising findings, why certain youth are susceptible, and some are not, to the disinhibiting effects of cyberspace and to engagement in aggressive online behavior remains relatively understudied.

Moral Disengagement and the Online Disinhibition Effect

A systematic review of Bandura's (2002) research into selective moral disengagement (i.e. a cognitive process by which a person justifies their own harmful or aggressive behavior towards others by loosening one's own inner self-regulatory mechanisms) (Bandura, 1986; 1999) posits that individuals experience greater ease towards engaging in harmful behavior when the harm is invisible to the perpetrator as a result of either distance or time. In support of Bandura's theoretical position, moral disengagement has frequently been linked to hostile, aggressive, and disinhibited behavior (Pornari & Wood, 2010; Runions & Bak, 2015). Extending these findings to the digital environment seems appropriate as this context provides

a felicitous landscape for moral disengagement and the expression of anti-social behaviors and attitudes that run counter to everyday norms (Bauman, 2009; Runions & Bak, 2015).

The behavioral juxtaposition outlined above, namely that people often behave differently online than in an offline context, has been described as the '*online disinhibition effect*' (Joinson, 1998; Joinson, 2003; Suler, 2004). According to Suler (2004), this effect can be further categorized as either '*benign*' or '*toxic*' digital disinhibition. In the case of benign disinhibition, the online environment motivates individuals to over-share personal details about themselves and their emotions. These individuals use the Internet as a means of exploring their inner self, and their over-sharing is marked by an intrinsic need to better understand existing or new emotions while working out interpersonal issues. In contrast, toxic disinhibition is characteristic of the modern troll, and is illustrated by displays of rude or crude language, harsh commentary, 'hate speech', and even threats that would be extremely rare in a face-to-face setting.

Following from this distinction between benign and toxic disinhibition, Suler (2004) suggested that there are several factors associated with the digital landscape which contribute to the online disinhibition effect, namely: dissociative anonymity (e.g., "They'll never know who I really am"), invisibility (e.g., "I can't see you, so you can't see me"), asynchronicity (e.g., "I'll post whatever I want now, and you'll see it later when I don't have to deal with your reaction"), solipsistic introjection (e.g., "The way I see you is the real you"), dissociative imagination (e.g., "Who I am online is different from who I am in real life"), and minimization of authority (e.g., "There

are no consequences for what I say or do online"). These six elements of online disinhibition promote self-disclosure through computer-mediated communication and digital technology. Supporting this proposition, a number of studies have found that behaviors of self-disclosure are significantly increased in the digital environment when compared to face-to-face interactions (Joinson, 2001; McKenna & Bargh, 1998; Tidwell & Walther, 2002).

Exploring Adolescent Identity and the Online False Self

The Internet provides adolescents with a unique space for exploring various facets of their identity and the freedom for dissociative self-expression (Bauman, 2010; Suler, 2004). In fact, some studies have shown that a large proportion of youth believe cyberspace provides them with a relatively safe and anonymous digital bubble that is free of the direct criticism, judgement, or immediate consequence that they may experience offline (Bauman, 2010; Runions & Bak, 2015; Suler, 2004). Moreover, the majority of teens confess to sharing a different self offline compared to the one they believe they share online (Kaplan & Haenlein, 2010). This belief system, coupled with the extensive interaction youth have with cyberspace, has important implications for the healthy development of their identity, and consequently their behavior. Adolescents' engagement with the Internet, for example, can impact not only the coherence, but also the integrity, of youth's developing sense of self.

An individual's self-concept, or true self, is achieved through a process of identity exploration (Marcia, 1966). For some youth, this is a difficult and arduous process that, if not carefully nurtured, can result in identity confusion. When

adolescents become confused about their identity, many develop a protective mask meant to diffuse the disparity between who one really is, and how they want others to perceive them (Dayton, 2011; Goth et al., 2012; Schwartz et al., 2011). These self-conceptions (i.e., one's relatively enduring and stable sense of 'the real me') can greatly influence the way adolescents represent themselves online (Gil-Or, Levi-Belz, & Turel, 2015).

To date, research has mainly focused on how perceptions of identity influence Facebook behavior and attitudes. For example, after measuring levels of 'true self' expression, Seidman (2014) found that higher levels of true self (authenticity) were associated with increased Facebook communication, self and emotional disclosure, as well as attention- and acceptance-seeking. A separate study found that adolescents who harboured higher false self perceptions, believed that information shared online is inaccurate, and not representative of real life (Kurek & Jose, 2016). Investigations into whether adolescent self-perceptions promote disinhibited behavior, however, have not been previously explored. The present study was designed to investigate whether underlying personality constructs may have a significant influence on the process of identity exploration and the formation of these personal self-perceptions. In particular, we sought to identify whether specific personality constructs prone to moral disengagement (i.e., aspects of aversive personality) may impact on identity and self-expression online.

The Dark Side of Personality and Cyberspace

Leading experts in the field of the 'dark personality' argue that there is a set of common, socially aversive traits known as the *dark tetrad* (Paulhus, 2014).

Individuals who express these personality traits of narcissism, psychopathy, Machiavellianism, and/or sadism, are seen to be self-centered and socially offensive, while often maintaining the ability to get along with other people in everyday settings, i.e., be considered as effectively engaging with the broader community (Paulhus, 2014). In the present study, the influence of three out of the four dark tetrad traits were explored: *narcissists,* who are often grandiose self-promoters, carry a strong sense of entitlement, lack empathy, and display high levels of egotism (Campbell & Miller, 2012; Madan, 2014; Paulhus, 2014); *psychopaths,* who cause others serious harm in impulsive fits of callous thrill-seeking, exhibit elevated selfishness, low inhibition, superficial charm, callousness, and display a lack of empathy or remorsefulness (Madan, 2014; Paulhus, 2014); and *sadists,* who try to verbally or physically hurt others for pleasure or amusement (Buckels, Jones, & Paulhus, 2013).

Recent work in cyber psychology has revealed that these dark personality characteristics contribute to not only social media network preferences but also to online behaviors (Buckels, Trapnell, & Paulhus, 2014; Cheng, Bernstein, Danescu-Niculescu-Mizil & Leskovec, 2017; Madan, 2014; van Geel, Goemans, Toprak, & Vedder, 2017). Frequently, narcissists display higher levels of activity on social media platforms and exhibit increased levels of self-promoting and self-enhancing behaviors (Aboujaoude, 2017; Choi, Panek, Nardis, & Toma, 2015; Halpern, Valenzuela, & Katz, 2016). Narcissists are also found to use social media networks for self-enhancement and to develop and exploit shallow and short-term online friendships in an attempt to bolster their social status and self-esteem (Barry,

Doucette, Loflin, Rivera-Hudson, & Herrington, 2017; Fox & Rooney, 2015; McCain et al., 2016).

Narcissism, however, is not the only dark personality to be manifested online. Both sadism and psychopathy have been linked to hostile, aggressive, and intimidating behaviors, which are frequently, characterized as trolling behavior in the online context (Buckels et al., 2014). Sest and March (2017) found that while higher levels of trait psychopathy and sadism predicted trolling behavior, individuals scoring highest on trait psychopathy tend to use trolling in an aim to manipulate others. Sadistic individuals, on the other hand, exhibit the highest levels of enjoyment of their own adversarial and provocative online behaviors (Buckels et al., 2014). This strong correlation between trolling and sadism illustrates how both trolls and sadists "feel sadistic glee at the distress of others" (Buckels et al., 2014, p. 101), suggesting that these individuals do it for the 'LULZ' (i.e., aggressive laughter derived from another person's distress or discomfort).

Recently a growing literature has begun to demonstrate that traits associated with narcissism, sadism, and psychopathy manifest in unhealthy digital behaviors (e.g., Grothe, Staar, & Janneck, 2016). However, little to no attention has been given to how these dark personality traits inform the expression of inauthentic identities, online disinhibition, or cyber aggressive behavior. The chief aim of the present study, therefore, is to describe the associations that likely exist between and among these various constructs.

Aims of the Current Study

While existing associations between each of the dark personality traits and false self have been previously established (e.g., Campbell & Foster, 2011; Vaknin, 2015), in the present study we endeavour to establish whether positive predictive associations between narcissism, sadism, and psychopathy exist in relation to online disinhibition, as well as cyber aggression. Moreover, the possible influence of adolescent false self perceptions on uninhibited digital behavior or cyber aggression will be explored, as the associations between these constructs are not well-established. Lastly, although links between increased online disinhibition and cyberbullying have been acknowledged (e.g., Udris, 2014), the present study aims to further investigate these effects. Following these distinctive gaps in the existing literature, the present study sought to investigate whether any of the dark personality traits would directly, or indirectly, predict false self perceptions, online disinhibition, and aggressive online behavior.

Firstly it is predicted that psychopathy, narcissism, and sadism will evidence significant and positive direct associations with perceptions of false self, online disinhibition, and cyber aggression (*H1*). Secondly, it is hypothesized that perceptions of false self will positively predict online disinhibition, while online disinhibition will be a significant positive predictor of cyber aggression (*H2*). Third, it is hypothesised that perceptions of false self will mediate the relationship between dark personality traits and online disinhibition, while online disinhibition will mediate the relationship between the dark traits and cyber aggression (*H3*). Finally, it is hypothesised that dark personality traits will have a significant effect on

cyber aggression through the double mediation of false self perceptions and online disinhibition (*H4*).

Methods

Participants

A total of 709 adolescents (50.5% female; 49.5% male) aged 13 to 17 years (M_{age} = 15.56 years) were recruited from 18 different high schools across both the North and South islands of New Zealand. Respondents reported on their cultural background, with a large proportion identifying as New Zealand European (67.1%; *n* = 476), the majority cultural group in New Zealand. Other ethnicities reported were Maori (16.9%; *n* = 120), Pacific Islander (2.2%; *n* = 16), Asian (10.3%; *n* = 74), and 3.3% (n = 24) identified as Other.

The key questions regarding online behavior and the dark personality traits are embedded in a larger battery of personality questionnaires and questions about information and communication technology (ICT) use. Participation was completely anonymous and voluntary. Parental consent and youth assent forms were collected for participants under the age of 16, while student consent forms were collected for those individuals 16 years of age and older prior to data collection. Ethical approval was obtained from the [HOME UNIVERSITY] Human Ethics Committee, and all schools, principals, parents, and adolescents consented to the study's procedures prior to data collection.

Procedure

Named the New Zealand Youth and Technology Use Project, a large-scale cross-sectional, subject variable data collection was performed using a robust youth

demographic throughout New Zealand. The project featured key items that assess technology behaviors, personality, attitudes, as well as digital habits and trends. At the beginning of the recruitment phase, a New Zealand national school zones dictionary was acquired, and a project directory of all secondary schools was created to facilitate school recruitment. Approximately 30 school principals were randomly selected as were approached by both phone call and email with a brief outline of the project details and goals communicated. At the conclusion of the recruitment phase, either the head of school or principal of 18 schools expressed interest in participating in the project. Thus, approximately 60% of approached schools agreed to participate. Participating classes within schools were selected at random by the school principals, and research information and consent forms were sent home for parental signature. The lead researcher and research assistant visited each participating school to carry out in-school assessments and oversee all project material dissemination over a period of five weeks in 2015. Participants at each school completed the online survey through the use of either a Samsung tablet provided by the research team, or a school computer via private web link, on the day of the assessment. All measures were randomly ordered at the individual level and administered using Qualtrics via a private webhost. Final data were collected from both urban and rural schools, and represented equal numbers of mixed-gender and single-gender institutions. Each participant received a mini chocolate bar in compensation for his or her time. Schools were also provided an anonymized, aggregated report on key findings at the school level on conclusion of the study.

Measures

Narcissism. The revised 32-item Pathological Narcissism Inventory (PNI; Pincus et al., 2009) was used to assess rates of overall narcissism (e.g., 'I find it easy to manipulate people', 'I am disappointed when people don't notice me') and responses were collected on a six-point scale ranging from 0 (*not at all like me*) to 5 (*very much like me*). Consistent with previous studies on adolescent narcissism, which evidenced a Cronbach's α of .93 (Lee-Rowland, Barry, Gillen, & Hansen, 2016), the current study obtained a Cronbach's alpha of .93.

Sadism. The Comprehensive Assessment of Sadistic Tendencies (CAST; Buckels & Paulhus, 2013) measure, containing a total of 13 items, was administered to participants. Responses were collected to assess the degree of overall sadistic tendencies among the youth sample by measuring vicarious sadism (e.g., 'I enjoy playing the villain in games and torturing other characters'; seven items) and direct verbal sadism (e.g., 'When making fun of someone, it is especially amusing if they realize what I'm doing'; six items). The five items assessing direct physical sadism were omitted from the study for ethical reasons. Each question was rated on a five-point scale from 1 (*strongly disagree*) to 5 (*strongly agree*). Previous work by Buckells et al. (2014), evidenced a complete scale reliability of α = .89, whereas the current study's total scale (with five omitted items) resulted in an acceptable Cronbach's alpha of .79.

Psychopathy. In order to measure the interpersonal and affective traits of psychopathy found to be persistent across development (Frick & White, 2008; White & Frick, 2010), the Inventory of Callous-Unemotional Traits (ICU; Frick, 2004)

was employed. A total of 12 items from the self-report questionnaire were used to assess the degree of *callousness* (e.g., 'I do not care who I hurt to get what I want'; seven items), and *uncaring* (e.g., 'I feel bad or guilty when I do something wrong'; five items all of which were reverse scored) in the sample. These items were scored on a three-point Likert scale from 1 (*not at all true*) to 3 (*definitely true*). The reliability and validity of the total ICU scores has been found to be consistently acceptable, with Cronbach's alphas ranging from .79 to .81 across various samples (Byrd, Kahn, & Pardini, 2014). With the removal of the unemotional subscale, the current study's total scale reliability resulted in an acceptable Cronbach's alpha of .71.

Machiavellianism. A revised 10-item version of the original 20-item self-report Kiddie Mach scale (Christie & Geis, 1970; Geng, Qin, Xia, & Ye, 2011) was used to assess Machiavellian traits and attitudes (e.g., Never tell anyone why you did something unless it will help you, anyone who completely trusts anyone else is asking for trouble, and sometimes you have to hurt other people to get what you want). Items were scored on a five-point Likert scale from 1 (disagree very much) and 5 (agree very much). Non-Machiavellian items were reverse-scored for consistency. The reliability of the total Kiddie-Mach scores has previously come under scrutiny for poor internal validity ranging from $\alpha = .29$ to .60 (Abell et al., 2015; Geng et al., 2011). Some researchers, however, have reported high internal validity ranging from $\alpha = .70$ to .79 (Andreou, 2004; Repacholi et al., 2003). The disparity in internal validity has led some researchers to caution against the use of the Kiddie-Mach when Cronbach's results are low, and in particular when they do

not reach the conventionally accepted α = .80 (Shea & Beatty, 1993). After adjusting the scale for brevity, the final 10-item scale resulted in a low Cronbach's alpha of .56. Due to previously reported research concerns about low reliability and validity of the scale, the modifications performed, and because the current measure did not achieve a level of reliability acceptable to testing, the Kiddie-Mach data obtained was omitted from the final analysis.

False Self. Perceptions of false self (POFS; Weir & Jose, 2010) were measured using 7 items (e.g., 'If people really knew what I was like on the inside, they wouldn't like me', 'I hide the real me by looking like others') scored on a five-point Likert scale from 1 (*strongly disagree*) to 5 (*strongly agree*). The scale's internal reliability has previously been reported with a Cronbach's α = .88, and 10-week test-retest reliability has been reported at r = .84 (Weir & Jose, 2010). The scale has also shown good convergent validity with other scale measures of false self (Say What I think Scale, Harter & Waters, 1991; and Silencing the Self Scale, Jack, 1991). The present scale resulted in a similarly high Cronbach's alpha of .86.

ICT attitudes and behaviors. For the purposes of the study the research team developed measures dedicated to assessing technology use and attitudes. This portion of the questionnaire asked participants to report on various ICT behaviors, preferences, motivations, and attitudes.

Online disinhibition. A total of 5 items assessing disinhibition online (e.g., 'I say/write/post comments online that would not say in person', 'I am aware when I am being hurtful online, but do it anyway', 'I feel safer expressing negative thoughts and feelings online than in person') were developed for the purpose of the study.

These items were scored on a 5-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The items selected loaded onto a single factor (see the psychometric evaluation of items below), and exhibited similar reliability to the 11 items used by Udris (2014). The total scale internal consistency was adequate in the current study (α = .73).

Aggressive online behavior. A total of 7 items assessing a range of aggressive online behaviors were developed for the purpose of this study. Participants were asked to rate how often in the last month (30 days) they had engaged in any of the listed behaviors (e.g., 'Posted something online about someone else to make others laugh', 'Made mean or negative comments on someone's photos, updates, or tags to make that person feel bad'). Items were scored on a 5-point Likert scale ranging from 1 (*Never*) to 5 (*7 or more times*). Higher scores indicate a higher frequency of aggressive behavior online and the total scale internal consistency was very strong (α = .92).

Psychometric Evaluation of Items

A total of twelve questions (see Appendix A) relating to both online disinhibition and cyber aggression were factor analyzed on a randomly selected half of the total sample using principal component analysis with Varimax (orthogonal) rotation. The eigenvalues of the first two factors, i.e., 4.93 and 2.16, as well as the scree plot, identified two distinct factors explaining a total of 59.16% of the variance of the entire set of variables. Factor 1 was labeled cyber aggression and explained 41.11% of the variance. The second factor derived was labeled online disinhibition

and explained 18.04% of the variance. All factor loadings were greater than .20 and there was a clear division between both factors.

The EFA-obtained factor structure was further examined with a CFA with the second randomly selected half of the sample. No cross-loadings were permitted in the CFA model, and it yielded good model fit indices (χ^2/df = 2.53, CFI = .97, RMSEA = .066; RMR = .048). These results provide support for the claim that we have identified two distinct factors of online behavior.

Data Analysis Plan

In order to assess the relationships between variables a preliminary investigation was first conducted through descriptive analyses and assessment of group differences by age and gender. In order to examine the substantive research hypotheses regarding the relationships among the aversive personality traits, false self perceptions, online disinhibition, and cyber aggression, a path model were constructed using AMOS.

Results

Descriptive Analysis

A missing values analysis was run on the sample data and results indicated that very little, .81%, data was missing. Little's MCAR test indicated that data were missing completely at random, $\chi^2 = 31.37$, df = 45, $p = .94$. In order to retain optimal statistical power, an expectation maximization imputation in SPSS was performed (Lin, 2010) and correlations, regressions, and MANOVA analyses were performed on this imputed dataset. The EM-imputed dataset was also employed for the SEM

analyses. Skewness and kurtosis estimates for all variables fell within the acceptable range, resulting in no data transformations.

Following the initial analysis, descriptive statistics were calculated. Table 1 presents the correlation coefficients, descriptive statistics, as well as Cronbach's *α*s for all of the variables in the study.

Table 1

Correlations among Dark Personality Traits and False Self, Online Disinhibition, and Aggressive Online Behavior

VARIABLES	NAR	SAD	PSY	POFS	ON-DIS	CYB-AGR
SAD	.25**	-				
PSY	.12**	.04	-			
POFS	.52**	16**	.16**	-		
ON-DIS	.41**	.31**	.01	.34**	-	
CYB-AGR	.17**	.26**	.05	.08*	.26**	
M	3.43	2.64	1.91	2.68	2.35	1.33
SD	.79	.68	.19	.61	.83	.62
α	.93	.79	.71	.86	.73	.92

Note. *$p < .05$, **$p < .01$. NAR = Narcissism, SAD = Sadism, PSY = Psychopathy, POFS = Perceptions of False Self, ON-DIS = Online Disinhibition, CYB-AGR= Cyber Aggression. (N = 718)

Gender and Age Differences

A MANOVA was conducted to determine whether the main variables in question significantly varied by the two covariates of gender and age. Age was split into two dichotomous groups of younger (13-14 years) and older (15-16 years)

adolescents. Significant main effects were discovered for both age, Wilk's Λ = .98, $F(6, 700) = 2.18$, $p < .05$, partial $\eta^2 = .02$, and gender, Wilk's Λ = . 76, $F(6, 700) = 37.09$, $p < .001$, partial $\eta^2 = .24$. An assessment of the univariate results revealed several significant group differences among the dependent variables (see Table 2). Females reported significantly higher levels of psychopathy, but lower levels of sadism, online disinhibition and cyber aggression than males Fs(1, 705) = 4.57 to 167.71, ps = .030 to .001, partial η^2s = .01 to .19. Age differences showed that younger adolescents reported lower narcissism, relative to older youth, $F(1, 705) = 7.61$, $p = .006$, partial $\eta^2 = .01$.

Table 2

Means and Standard Errors for Mean Group Comparisons of Gender and Age

	Males	Females	Younger	Older
Narcissism	3.38 (.04)	3.49 (.04)	3.36** (.04)	3.52 (.04)
Sadism	2.95*** (.03)	2.54 (.03)	2.65 (.03)	2.60 (.03)
Psychopathy	1.89** (.01)	1.93 (.01)	1.90 (.01)	1.92 (.01)
POFS	2.85 (.03)	2.88 (.03)	2.84 (.03)	2.89 (.03)
ON-DIS	2.45** (.04)	2.26 (.04)	2.34 (.04)	2.37 (.05)
CYB-AGR	1.38* (.03)	1.28 (.03)	1.36 (.03)	1.3 (.03)

Note. Younger = 13 -15 yrs; Older = 15 -17 yrs. Standard errors are presented within parentheses. *$p < .05$; **$p < .01$; ***$p < .001$. POFS = Perceptions of False Self, ON-DIS = Online Disinhibition, CYB-AGR= Cyber Aggression.

A Model of the Dark Personality Traits as Predictors of Online Behavior

In the interest of testing how the dark personality variables might work together to predict both online disinhibition and cyber aggression, a path model l examining all possible contiguous and indirect effects between variables was created. By conceptualizing the three dark personality traits as exogenous variables, it was posited that these are relatively stable characteristics of the person which predict down-stream context-specific and changeable characteristics and behaviors, i.e., false self perceptions, online disinhibition, and aggressive online behavior. Specifically, the model assessed direct effects from the dark personality traits (i.e., narcissism, sadism, and psychopathy) to false self perceptions, and also to online behaviors (i.e., online disinhibition and cyber aggression), in order to address hypotheses 1 and 2. The possible indirect effects of dark personality traits on online behaviors through adolescent perceptions of false self and disinhibition (see Figure 1) were also tested within the same model in order to address hypotheses 3 and 4.

The path model was fully saturated (degrees of freedom = 0) and allowed all possible paths to be freely estimated. Indirect effects were estimated with 5000 bootstrapped iterations, statistically evaluated with a bias corrected 95% confidence interval, and the *p*-values were calculated with a Monte Carlo estimates. Not shown in Figure 1 but estimated in the model were covariances among the three dark personality traits.

Exploring the Direct Effects among Variables

In the first step of the analyses, an initial examination of direct relationships between variables noted in the path model was carried out in order to test hypotheses 1 and 2. A significant relationship between the three dark personality

traits (e.g., narcissism, sadism, and psychopathy) and false self perceptions was hypothesised (*H1*). In partial support of the hypothesis, only narcissism (β = .38, SE = .026, p < .001) and psychopathy (β = .34, SE = .101, p < .001) evidenced positive significant effects, while sadism showed no significant direct effect ((β = .02, SE = .029, p = .41). Furthermore, the path model indicated that narcissistic traits (β = .22, SE = .053, p < .001) and sadism (β = .27, SE = .041, p < .001) were positive predictors of online disinhibition. Interestingly, and contrary to study predictions, the opposite effect was discovered for psychopathy (β = -.31, SE = .143, p = .03), which was found to be a marginal, but negative predictor of disinhibition online. It was also predicted that perceptions of false self would be a significant positive predictor of online disinhibition (*H2*), and this was supported empirically (β = .23, SE = .053, p < .001).

Next, it was hypothesized that dark personality (H1) and online disinhibition would be a significant predictor of aggressive online behavior (H2). Contrary to study predictions, only sadistic tendencies (β = .17, SE = .034, p < .001) were found to be a significant positive predictor of aggressive online behavior. Narcissism (β = .05, SE = .035, p = .151) and psychopathy (β = -.004, SE = .117, p = .971) showed no significant relationship. Instead, their impact on online aggression was indirect (see the mediation analyses below). However, supporting H2 it was found that online disinhibition significantly predicted aggressive online (β = .14, SE = .031, p < .001).

Mediation Analyses

In order to examine hypotheses 3 and 4, a fully saturated mediation analysis was conducted to assess if false self perceptions and online disinhibition

individually (single) or sequentially (double) mediated the relationship between the dark personality traits and digital behaviors (e.g., online disinhibition or cyber aggression). The mediation analyses revealed that out of a total of 13 possible indirect effects, 8 indirect effects yielded statistical significance. The results of these mediation analyses are reported in the subsequent section. Table 3 reports the results of the single mediation analyses, while Table 4 reports the results for the double mediations that were performed.

Online Disinhibition as the outcome variable. Using online disinhibition (ON-DIS) as the outcome variable, a mediation analysis was performed with narcissism, sadism, and psychopathy as the independent variables, and perceptions of false self as the mediator. In an aim to test *H3*, namely whether perceptions of false self would mediate the relationship between the dark personality and online disinhibition, this analysis revealed two statistically significant positive indirect effects (see Table 3), the first between narcissism ($\beta = .08$, $p < .001$) and online disinhibition, and the second, between psychopathy ($\beta = .08$, $p < .001$) and online disinhibition. Contrary to predictions, no statistically significant mediation was found between sadism and online disinhibition ($\beta = .01$, $p = .32$).

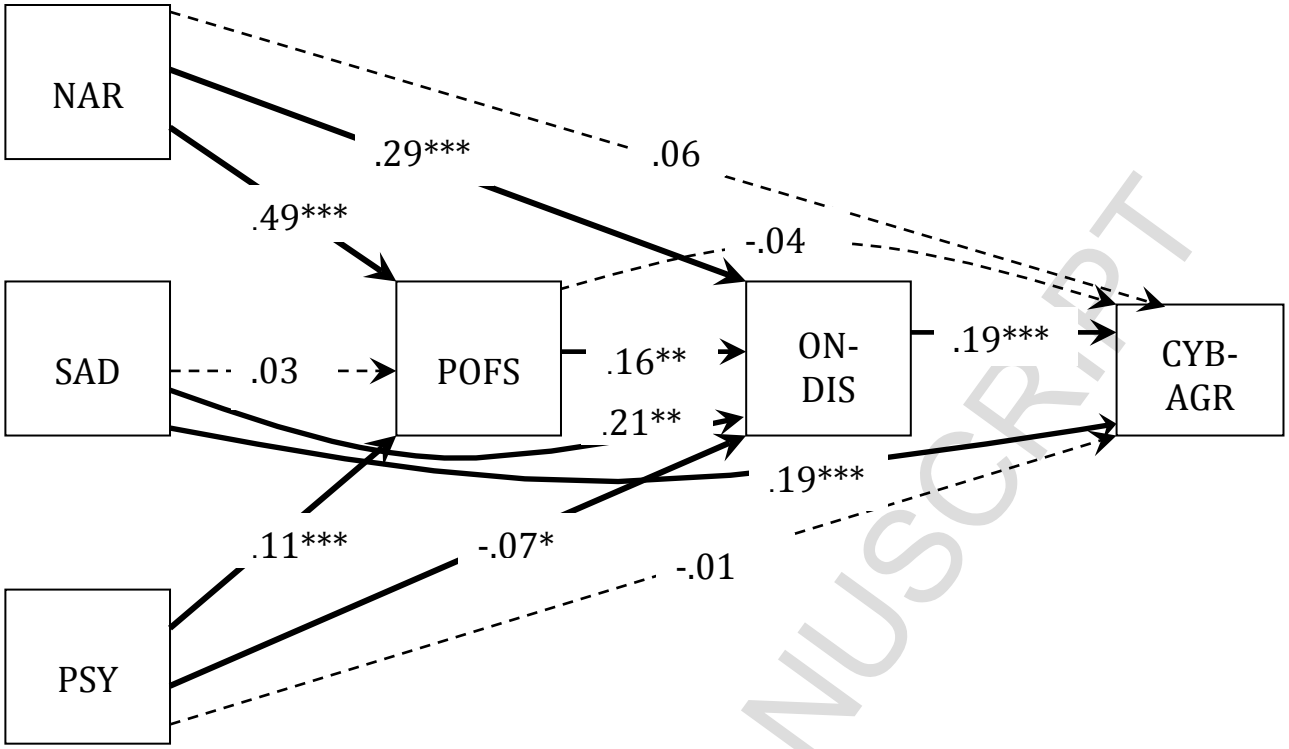

Figure 1. Model of the Dark Personality Traits as Predictors of False Self and Online Behavior. NAR = Narcissism; SAD = Sadism; PSY = Psychopathy; POFS = Perceptions of False Self; ON-DIS = Online disinhibition; CYB-AGR = Cyber Aggression. N = 709; *p < .05; **p < .01; ***p < .001; Solid lines indicate statistical significance whereas dashed lines are non-significant.

Cyber aggression as the outcome variable. Next, using cyber aggression (CYB-AGR) as the outcome variable, both perceptions of false self and online disinhibition were tested individually as mediating variables between the dark personality traits and CYB-AGR. Upon examining perceptions of false self as the mediating variable between the three independent variables and cyber aggression, no significant mediation was determined for narcissism (β = -.02, p = .26), sadism (β = -.001, p = .26), or psychopathy (β = -.02, p = .19). This result suggests that the dark personalities were not associated with increased rates of cyber aggression via

greater false self perceptions. In the second instance, and in line with *H3*, statistically significant mediations for narcissism (β = .04, p < .001), sadism (β = .04, p < .001), and psychopathy (β = -.04, p < .05) through online disinhibition to cyber aggression were obtained. This mediation result suggests that individuals who scored higher on all three aversive personalities were more likely to report higher levels of online disinhibition, which, in turn, predicted higher levels of cyber aggression. It is notable that all three dark personality traits contributed significant amounts of unique variance in this respect.

Table 3

Indirect Effects for Perceptions of False Self as a Mediating Variable.

Independent Variable	Mediating Variable	Dependent Variable	Indirect Effects	Standard Error	Ratio %	95% CI	Sig.
Narcissism	False Self	On-Dis	.084	.021	21.88%	[.046, .126]	.001
	False Self	Cyb-Agr	-.018	.016	25.93%	[-.051, .013]	.26
	On-Dis	Cyb-Agr	.042	.011	46.81%	[.024, .067]	.001
Sadism	False Self	On-Dis	.006	.007	2.14%	[-.006, .021]	.32
	False Self	Cyb-Agr	-.001	.002	.78%	[-.009, .001]	.26
	On-Dis	Cyb-Agr	.038	.10	20.30%	[.021, .061]	.001
Psychopathy	False Self	On-Dis	.079	.029	20.09%	[.033, .152]	.001
	False Self	Cyb-Agr	-.017	.016	9.91%	[-.056, .010]	.19
	On-Dis	Cyb-Agr	-.044	.023	23.55%	[-.100, -.007]	.02
False Self	On-Dis	Cyb-Agr	.031	.010	41.31%	[.015, .055]	.001

Note. On-Dis = Online Disinhibition; Cyb-Agr = Cyber Aggression; Ratio = effect size determined by indirect/direct.

Last, in order to better understand the predictors of cyber aggression, both perceptions of false self and online disinhibition were tested as sequential mediating variables. By introducing the mediators sequentially, the study hoped to

illuminate the possible mediating effects that may exist between elements of the dark personality and the outcome variable of cyber aggression (*H4*). Upon investigation, the results indicated that the double mediation (see Table 4) was statistically supported for both narcissism (β = .01, p < .001) and psychopathy (β = .01, p < .001), but not for sadism (β = .001, p = .29). These double mediation results suggest that adolescents who scored higher on either narcissism or sadism were more likely to experience increased perceptions of false self and online disinhibition, and finally to engage in more cyber aggressive behavior.

Table 4

Double Mediation of Inauthenticity and Online Disinhibition of the Relationship between Dark Personality Traits and Cyber Aggression

IndVar	MedV1	MedV2	DepVar	Indirect Effects	Standard Error	Ratio %	95% CI	Sig.
NAR	False Self	ON-DIS	CYB-AGR	.012	.004	5.80	[.006, .021]	.001
SAD	False Self	ON-DIS	CYB-AGR	.001	.001	.54	[-.001, .003]	.29
PSY	False Self	ON-DIS	CYB-AGR	.011	.005	5.73	[.004, .024]	.001

Note. IndVar = Independent Variable; DepVar = Dependent Variable; MedV1 = Mediating Variable 1; MedV2 = Mediating Variable 2; NAR = Narcissism; SAD = Sadism; PSY = Psychopathy; ON-DIS = Online Disinhibition; CYB-AGR = Cyber Aggression

Discussion

As societies become increasingly dependent on cyber technology, there is a pressing need for researchers to explore how facets of identity manifest in cyberspace behavior, particularly during adolescence. The chief aims of this study were to examine the influence of narcissism, sadism, and psychopathy in predicting online disinhibition and cyber aggression. The study also endeavoured to

investigate how perceptions of false self (i.e., identity inauthenticity) may mediate the relationship between the dark personality traits and online disinhibition. And last, the study sought to identify whether both perceptions of false self and online disinhibition might mediate the relationship between the dark personality traits and cyber aggression. Several significant results were found, namely that all three dark personality traits, as well as adolescent false self perceptions, were significantly and positively associated with increased online disinhibition. In addition, while only sadistic traits and online disinhibition were found to be significant direct predictors of cyber aggression, several indirect effects were also discovered, namely that all three dark traits became predictive of cyber aggression through the indirect role of increased disinhibition. Additionally, both narcissistic and psychopathic tendencies indirectly predicted cyber aggression through the mediation of both false self perceptions and online disinhibition.

Predictors of Online Disinhibition

In the development of the online disinhibition effect theory, Suler (2004) proposed that the concept of self as well as underlying feelings, needs, drives, and personality variables, have the potential to foster and promote the online disinhibition effect. The results of the present study are the first to provide empirical evidence that suggests that some of these latent mechanisms may indeed significantly predict an increase in the expression of online disinhibition. The findings revealed that both narcissism and sadism positively predicted online disinhibition. Higher scores on perceptions of false self (i.e., inauthenticity) also significantly predicted online disinhibition. The opposite effect was discovered for

psychopathy, namely that adolescents who reported high levels of psychopathy exhibited *decreased* online disinhibition. A possible explanation for this effect could be found in the theory of context-dependent psychopathic behavior that posits that psychopathic expression can differ in various circumstances, including possibly the online and offline worlds (Koenigs, Baskin-Sommers, Zeier, & Newman, 2011; Nevin, 2015; Vaknin, 2015). At the core of the psychopathic personality are traits of low conscientiousness, indifference to the opinion of others, proneness to boredom, and impulsivity (Freeman, 2014); it is possible that these traits may not be amplified by the disinhibiting nature of cyberspace. In short, some psychopathic individuals may not experience any disinhibiting effect of being online; their baseline of disinhibition may be chronically high.

Delving deeper into these associations, the construct of false self perceptions was tested as a mediating variable between each of the dark personality traits and online disinhibition. It was found that false self perceptions play a significant role in mediating the relationship between narcissism and online disinhibition, as well as psychopathy and online disinhibition, indicating that an increase in either of these personality traits results in an increase of false self perceptions and, in turn, predicts an increase in online disinhibition. These findings indicate that the long-established link in adults between narcissism and false self (Vaknin, 2015) might arise as early as adolescence. It was also found that youth manifesting psychopathic tendencies are likely to report higher levels of inauthentic self, which, in turn, was found to be predictive of higher disinhibition. This pattern suggests that psychopathic youth experiencing heightened false self perceptions may be more inclined towards

behaving in a more impulsive and disinhibited fashion online, potentially as a defense mechanism or in retaliation when their fragile or false belief systems or opinions are challenged. Alternatively, perhaps these youth are still in the process of identity exploration, and are struggling to reconcile their psychopathic inclinations in the offline world, so they turn to the Internet to exhibit suppressed psychopathic urges.

These results suggest, as Suler (2004) noted, that the underlying mechanisms of inhibited or disinhibited behavior lie fundamentally within the processes of personality dynamics, and consistent with this view, the present findings provide empirical evidence that certain individuals may be at an increased risk of disinhibited behavior online. The developmental trajectory from childhood to young adulthood is often affected by growing pains of risk behavior, impulsivity, tendencies towards misconduct, and egocentrism, all of which are associated with traits of psychopathy and both adaptive and maladaptive narcissism (Chinchilla & Kosson, 2016; Hill & Lapsley, 2011; van den Bos & Hertwig, 2017). Perhaps certain adolescents, due to a vulnerability created by the presence of dark personality traits, are more susceptible to the unique nature of cyberspace and certain digital platforms that stimulate and encourage online behavior that is less inhibited compared to offline behavior.

For example, Aboujade (2011) argues that the nature of the Internet facilitates easy self-promotion and instant gratification of individual needs. Some youth, whose basic needs are grounded in dark motives, when exposed to a vastly uncontrolled and unmonitored space like the Internet, may gravitate towards the

unhealthy development of disinhibited actions and attitudes. Results of the present study provide validation for Suler's (2004) argument that certain personality traits predispose individuals to engage in more online disinhibition. This basic foundation can be easily aggravated by external factors such as environmental or societal influences. For example, Casale, Fiovaranti, and Caplan (2015) found that unique features of computer-mediated communication, such as a reduction in nonverbal cues and enhanced message control, brings on greater feelings of disinhibition and increased preference for online social interactions among adolescents. This suggests that the digital world does, to some degree, influence digital behavior.

Predictors of Cyber Aggression

Cyberspace has become an instrument for both positive and negative social influence. Trolling behavior, in particular, has garnered much attention in recent years. In an effort to understand the motives that trigger the distasteful and antisocial activities of those who engage in aggressive behaviors online, the present investigation yielded several novel and important results in predicting cyber aggression. The investigation of particular pathways through mediation revealed additional unique associations among the variables. Namely, it was found that all three dark traits, false self perceptions, and online disinhibition predicted, either directly or indirectly, greater cyber aggression. Sadism, for example, unlike narcissism and psychopathy, was the only personality construct found to be directly predictive of cyber aggression, indicating that sadism does not have an influence on cyber aggression through false self, while narcissism and psychopathy both do. Looking at these associations more closely, it could be said that sadistic individuals

may hold a unique baseline motivation for engaging in trolling behavior online that is independent of these false self perceptions. Specifically, while these false self perceptions have been found to play a significant role in predicting online disinhibition, they have evidenced no such role in directly predicting cyber aggression. This suggests that of the dark personality traits, it is the sadistic youth who engage in cyber aggression that may be more intrinsically motivated, and find particular pleasure or amusement, in trolling behavior (Buckels et al., 2013; Levesque, 2011). This is broadly consistent with past research that has suggested that a common adolescent motive behind aggressive online behavior is humour and doing it 'for fun' (Bartlett et al., 2014; Kyriacou, 2015), traits consistently associated with characteristics of sadism.

Lastly, it was identified that both narcissism and psychopathy significantly predicted greater perceptions of false self and higher levels of online disinhibition that resulted in predictive elevated engagement in cyber aggressive behavior. In the case of elevated false self perceptions in conjunction with psychopathic traits, it is possible that the characteristic impulsivity frequently recognized as a common trait of psychopathy may be exacerbated in digital settings when the fragile self-concept of these individuals is tested, resulting in an increased response to perceived provoked attacks or through the disinhibition of social restraints (Fanti, Frick, & Georgiou, 2009; Madan, 2014). Psychopathy has also been positively associated with trolling behaviors, specifically of social media profiles of popular or prominent users (Lopes & Yu, 2017), illustrating a somewhat retaliatory or exploitative motive behind the cyber aggressive activity. The double mediation outcomes found point to

how several steps may be needed to explain the ways in which personality is predictive of cyber aggression. Taken together, these findings are consistent with, and further explicate, previous work by having identified mechanisms of how the dark personality predicts cyber aggression through various mediators.

Future Research and Limitations

Future research should examine in a more detailed fashion the linkages identified here and elsewhere. In the case of narcissistic individuals, who desire admiration and see themselves as superior to others (Lopes & Yu, 2017), research should examine whether cyber aggressive acts committed by this demographic are more deliberate, manipulative, and instrumental in nature (Madan, 2014; Wright, 2017) than in the case of psychopathic youth, who have a more impulsive nature (Lopes & Yu, 2017). In addition, research should aim to investigate how the disinhibiting characteristics of the cyberspace environment may exacerbate trolling activities and aggressive behavior among these groups of adolescents. Additionally, research into possible disparities between offline and online self-views may shed light on why cyberspace may widen the gap between these two dichotomous selves and behavior. Consideration should be given to how the perceptions of a safe space in the Internet may facilitate individuals harbouring dark personality traits to act in disinhibited and aggressive ways to help them bolster their fragile and evolving self-concepts and self-image vulnerabilities (Barry, Kerig, Stellwagen, & Barry, 2011).

The main limitation of the study is that the data used were concurrent in nature, which prevented us from drawing firm conclusions about sequential and temporal influences. Furthermore, the use of longitudinal data may prove to

increase the reliability of the Kiddie-Mach scale (Geng et al., 2011). Moreover, while the distinctiveness between the Dark Triad traits in non-clinical adolescent samples has been disputed (McHoskey, 2001) the inclusion of the Kiddie Mach in future research may provide a more holistic picture of the effects of the dark triad on online disinhibition and aggressive online behavior. It is also noteworthy that, with the ubiquitous online presence of youth and adults a like, further work is needed in the development of an online disinhibition measure to advance both the validity and reliability of the construct.

Lastly, having only measured sub-clinical rather than clinical levels of the dark personality traits, the present findings may not generalize to clinical samples. Third, the sample, while relatively large in size, solely represents New Zealand high school students and thus may not be representative of other age populations or cultures.

Conclusions

In sum, this study revealed important associations between three dark personality traits, false self perceptions, online disinhibition, and cyber aggression. Specifically, all of these variables were found to be positively associated with each other in both direct and indirect ways. These findings demonstrate the importance of delving deeper into the reasons why individuals engage in disinhibited and cyber aggressive behavior online. In an increasingly cyber-connected world that is progressively utilized by children and adolescents, there is a growing urgency to understand how the nature of the digital world disinhibits negative behavior, and in particular, how certain maladaptive personality traits trigger dangerous and hurtful

online behavior. We hope that the present findings provide novel and informative implications so that interventions can be created to reduce the incidence of antisocial, disinhibited, and aggressive digital behavior.

References

Abell, L., Qualter, P., Brewer, G., Barlow, A., Stylianou, M., Henzi, P., & Barrett, L. (2015). Why Machiavellianism Matters in Childhood: The Relationship Between Children's Machiavellian Traits and Their Peer Interactions in a Natural Setting. *Europe's Journal of Psychology*, *11*(3), 484-493. doi:10.5964/ejop.v11i3.957

Aboujaoude, E. (2011). *Virtually you: The dangerous powers of the e-personality*. New York, NY: W.W. Norton.

Aboujaoude, E. (2017). The Internet's effect on personality traits: An important casualty of the "Internet addiction" paradigm. *Journal of Behavioral Addictions*, *6*(1), 1-4. doi:10.1556/2006.6.2017.009

Anderson, C. A., & Bushman, B. J. (1001). Human aggression. *Annu. Rev. Psychol*, *53*, 27-51. doi:10.1146/annurev.psych.53.100901.135231

Andreou, E. (2004). Bully/victim problems and their association with Machiavellianism and self-efficacy in Greek primary school children. *British Journal of Educational Psychology*, *74*(2), 297-309. doi:10.1348/000709904773839897

Bandura, A. (1986*). Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review,* 3, 193-209.

Bandura, A. (2002). Social cognitive theory in cultural context. *Applied Psychology*, *51*(2), 269-290. doi:10.1111/1464-0597.00092

Barlett, C. P., Gentile, D. A., Anderson, C. A., Suzuki, K., Sakamoto, A., Yamaoka, A., & Katsura, R. (2013). Cross-cultural differences in cyberbullying behavior. *Journal of Cross-Cultural Psychology*, *45*(2), 300-313. doi:10.1177/0022022113504622

Barry, C. T. (2011). *Narcissism and Machiavellianism in youth: Implications for the development of adaptive and maladaptive behavior*. Washington, DC: American Psychological Association.

Barry, C. T., Doucette, H., Loflin, D. C., Rivera-Hudson, N., & Herrington, L. L. (2017). "Let me take a selfie": Associations between self-photography, narcissism, and self-esteem. *Psychology of Popular Media Culture*, *6*(1), 48-60. doi:10.1037/ppm0000089

Bauman, S. (2009). Cyberbullying in a rural intermediate school: An exploratory study. *The Journal of Early Adolescence*, *30*(6), 803-833. doi:10.1177/0272431609350927

Buckels, E. E., Jones, D. N., & Paulhus, D. L. (2013). Behavioral confirmation of everyday sadism. *Psychological Science*, *24*(11), 2201-2209. doi:10.1177/0956797613490749

Buckels, E. E., & Paulhus, D. L. (2014). *Comprehensive Assessment of Sadistic Tendencies* (CAST). Unpublished instrument, University of British Columbia, Vancouver, Canada.

Buckels, E. E., Trapnell, P. D., & Paulhus, D. L. (2014). Trolls just want to have fun. *Personality and Individual Differences*, *67*, 97-102. doi:10.1016/j.paid.2014.01.016

Byrd, A. L., Kahn, R. E., & Pardini, D. A. (2013). A validation of the Inventory of Callous-Unemotional Traits in a community sample of young adult males. *Journal of Psychopathology and Behavioral Assessment*, *35*(1), 20-34. doi:10.1007/s10862-012-9315-4

Campbell, W. K., & Foster, J. D. (2011). The narcissistic self: Background, an extended agency model, and ongoing controversies. In C. Sedikides & S. J. Spencer (Eds.), *The self: Frontiers of social psychology*. New York, NY: Psychology Press.

Campbell, W. K., & Miller, J. D. (2012). *The handbook of narcissism and narcissistic personality disorder: Theoretical approaches, empirical findings, and treatments*. doi:10.1002/9781118093108

Casale, S., Fiovaranti, G., & Caplan, S. (2015). Online disinhibition: Precursors and outcomes. *Journal of Media Psychology: Theories, Methods, and Applications*, *27*(4), 170-177. doi:10.1027/1864-1105/a000136

Cheng, J., Bernstein, M., Danescu-Niculescu-Mizil, C., & Leskovec, J. (2017). Anyone can become a troll. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing - CSCW '17*, 1217-1230. doi:10.1145/2998181.2998213

Chinchilla, M. A., & Kosson, D. S. (2016). Psychopathic Traits Moderate Relationships Between Parental Warmth and Adolescent Antisocial and Other High-Risk Behaviors. *Criminal Justice and Behavior*, *43*(6), 722-738. doi:10.1177/0093854815617216

Choi, M., Panek, E. T., Nardis, Y., & Toma, C. L. (2015). When social media isn't social: Friends' responsiveness to narcissists on Facebook. *Personality and Individual Differences*, *77*, 209-214. doi:10.1016/j.paid.2014.12.056

Christie, R., & Geis, F. L. (1970). *Studies in Machiavellianism*. New York: Academic Press.

Corcoran, L., Guckin, C., & Prentice, G. (2015). Cyberbullying or cyber aggression?: A review of existing definitions of cyber-based peer-to-peer aggression. *Societies*, *5*(2), 245-255. doi:10.3390/soc5020245

Crocker, J., & Wolfe, C. T. (2001). Contingencies of self-worth. *Psychological Review, 108(3),* 593-623. doi: 10.1037/0033-295X.108.3.593

Dayton, T. (2011, November 17). Creating a false self: Learning to live a lie [Web log post]. Retrieved from https://www.huffingtonpost.com/dr-tian-dayton/creating-a-false-self-lea_b_269096.html

Fanti, K. A., Frick, P. J., & Georgiou, S. (2009). Linking callous-unemotional traits to instrumental and non-instrumental forms of aggression. *Journal of Psychopathology and Behavioral Assessment*, *31*(4), 285-298. doi:10.1007/s10862-008-9111-3

Fox, J., & Rooney, M. C. (2015). The Dark Triad and trait self-objectification as predictors of men's use and self-presentation behaviors on social networking sites. *Personality and Individual Differences*, *76*, 161-165. doi:10.1016/j.paid.2014.12.017

Freeman, R. (2014). Basic differences between psychopathy & narcissistic personality disorder. Retrieved from https://neuroinstincts.com/a-few-

basic-differences-between-psychopathy-narcissistic-personality-disorder-part-one/

Frick, P. J. (2004). *Inventory of callous–unemotional traits.* Unpublished rating scale, University of New Orleans, New Orleans, LA.

Frick, P. J., & Ellis, M. (1999). Callous-unemotional traits and subtypes of conduct disorder. *Clin Child Fam Psychol Rev.*, *2*(3), 149-168.

Garett, R., Lord, L. R., & Young, S. D. (2016). Associations between social media and cyberbullying: A review of the literature. *mHealth*, *2*, 46-46. doi:10.21037/mhealth.2016.12.01

Geng, Y., Qin, B., Xia, D., & Ye, Q. (2011). Reliability and Validity of the Kiddie Mach Scale in Chinese Children. *Psychological Reports*, *108*(1), 229-238. doi:10.2466/03.09.17.pr0.108.1.229-238

Gil-Or, O., Levi-Belz, Y., & Turel, O. (2015). The "Facebook-self"□: Characteristics and psychological predictors of false self-presentation on Facebook. *Frontiers in Psychology*, *6*, 99. doi:10.3389/fpsyg.2015.00099

Goth, K., Foelsch, P., Schlüter-Müller, S., Birkhölzer, M., Jung, E., Pick, O., & Schmeck, K. (2012). Assessment of identity development and identity diffusion in adolescence - Theoretical basis and psychometric properties of the self-report questionnaire AIDA. *Child and Adolescent Psychiatry and Mental Health*, *6*(1), 27. doi:10.1186/1753-2000-6-27

Grothe, M., Staar, H., & Janneck, M. (2016, June). *How to treat the troll? An empirical analysis of counterproductive online behavior, personality traits and organizational behavior*. Paper presented at 11th International Forum on

Knowledge Asset Dynamics, Dresden. Retrieved from https://www.researchgate.net/publication/305208082_How_to_treat_the_troll_An_empirical_analysis_of_counterproductive_online_behavior_personality_traits_and_organizational_behavior

Halpern, D., Valenzuela, S., & Katz, J. E. (2016). "Selfie-ists" or "Narci-selfiers"?: A cross-lagged panel analysis of selfie taking and narcissism. *Personality and Individual Differences*, *97*, 98-101. doi:10.1016/j.paid.2016.03.019

Harter, S. & Waters, P. L. (1991). Saying What I Think Around Others. Unpublished manuscript. Denver, CO; University of Denver

Hemphill, S. A., & Heerde, J. A. (2014). Adolescent predictors of young adult cyberbullying perpetration and victimization among Australian youth. *Journal of Adolescent Health*, *55*(4), 580-587. doi:10.1016/j.jadohealth.2014.04.014

Hill, P. L., & Lapsley, D. K. (2011). Adaptive and maladaptive narcissism in adolescent development. In C. T. Barry, P. Kerig, K. Stellwagen, & T. D. Barry (Eds.), *Implications of narcissism and Machiavellianism for the development of prosocial and antisocial behavior in youth* (pp. 89-105). Washington, DC: American Psychological Association. doi:10.1037/12352-005

Hymel, S., Schonert-Reichl, K. A., Bonanno, R. A., Vaillancourt, T., & Henderson, N. R.(2010). Bullying and morality: Understanding how good kids can behave badly. In S. R. Jimerson, S. M. Swearer, & D. Espelage (Eds.), *The handbook of school bullying: An international perspective*. New York, NY: Routledge.

Hymel, S., & Bonanno, R. A. (2014). Moral disengagement processes in bullying. *Theory Into Practice*, *53*(4), 278-285. doi:10.1080/00405841.2014.947219

Jack, D. C. (1991). Silencing the Self: Women and Depression. Cambridge: Harvard University Press.

Joinson, A. N. (1998). Causes and effects of disinhibition on the Internet. In J. Gackenbach (Ed.) *The psychology of the Internet* (pp. 43-60). New York. Academic Press.

Joinson, A. N. (2001). Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity. *European Journal of Social Psychology, 31*, 177-192.

Joinson, A. N. (2003). *Understanding the psychology of Internet behaviour: Virtual worlds, real lives.* Basingstoke and New York: Palgrave Macmillan.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. *Business Horizons*, *53*(1), 59-68. doi:10.1016/j.bushor.2009.09.003

Koenigs, M., Baskin-Sommers, A., Zeier, J., & Newman, J. P. (2010). Investigating the neural correlates of psychopathy: a critical review. *Molecular Psychiatry*, *16*(8), 792-799. doi:10.1038/mp.2010.124

Kowalski, R. M., Giumetti, G. W., Schroeder, A. N., & Lattanner, M. R. (2014). Bullying in the digital age: A critical review and meta-analysis of cyberbullying research among youth. *Psychological Bulletin*, *140*(4), 1073-1137. doi:10.1037/a0035618

Kurek, A. M., & Jose, P. E. (2016, April). *Can self-perceptions predict Facebook behaviours & beliefs?* Paper presented at The 2016 Society for Research on Adolescence Biennial Meeting, Baltimore.

Kyriacou, C. (2015, October). *A psychological typology of cyberbullies in schools*. Paper presented at British Psychological Society Psychology of Education Section Conference, Liverpool.

Lee-Rowland, L. M., Barry, C. T., Gillen, C. T., & Hansen, L. K. (2016). How do different dimensions of adolescent narcissism impact the relation between callous-unemotional traits and self-reported aggression? *Aggressive Behavior*, *43*(1), 14-25. doi:10.1002/ab.21658

Levesque, R. J. (2011*).* Sadistic Personality Disorder*. Encyclopedia of Adolescence*. (pp. 2445). doi: 10.1007/978-1-4419-1695-2.

Lin, T. H. (2010). A comparison of multiple imputation with EM algorithm and MCMC method for quality of life missing data. *Quality & Quantity*, *44*(2), 277-287. doi:10.1007/s11135-008-9196-5

Lopes, B., & Yu, H. (2017). Who do you troll and why: An investigation into the relationship between the dark triad personalities and online trolling behaviours towards popular and less popular Facebook profiles. *Computers in Human Behavior*, *77*, 69-76. doi:10.1016/j.chb.2017.08.036

Madan, A. O. (2014). Cyber aggression/cyber bullying and the Dark Triad: Effect on workplace behavior/performance. *International Journal of Computer and Systems Engineering*, *8*(6), 1740-1745. Retrieved from

https://waset.org/publications/9998533/cyber-aggression-cyber-bullying-and-the-dark-triad-effect-on-workplace-behavior-performance

Malti, T., Gasser, L., & Gutzwiller-Helfenfinger, E. (2010). Children's interpretive understanding, moral judgments, and emotion attributions: Relations to social behaviour. *British Journal of Developmental Psychology*, *28*(2), 275-292. doi:10.1348/026151009x403838

Marcia, J. E. (1966). Development and validation of ego-identity status. *Journal of Personality and Social Psychology*, *3*(5), 551-558. doi:10.1037/h0023281

McCain, J. L., Borg, Z. G., Rothenberg, A. H., Churillo, K. M., Weiler, P., & Campbell, W. K. (2016). Personality and selfies: Narcissism and the Dark Triad. *Computers in Human Behavior*, *64*, 126-133. doi:10.1016/j.chb.2016.06.050

McKenna, K. Y., & Bargh, J. A. (1998). Coming out in the age of the Internet: Identity "demarginalization" through virtual group participation. *Journal of Personality and Social Psychology*, *75*(3), 681-694. doi:10.1037//0022-3514.75.3.681

Menesini, E., Nocentini, A., & Camodeca, M. (2011). Morality, values, traditional bullying, and cyberbullying in adolescence. *British Journal of Developmental Psychology*, *31*(1), 1-14. doi:10.1111/j.2044-835x.2011.02066.x

Menesini, E., Nocentini, A., & Camodeca, M. (2013). Morality, values, traditional bullying, and cyberbullying in adolescence. *British Journal of Developmental Psychology*, *31*(1), 1-14. doi:10.1111/j.2044-835x.2011.02066.x

Nevin, Andrew D. (2015). Cyber-Psychopathy: Examining the Relationship between Dark E-Personality and Online Misconduct. *Electronic Thesis and Dissertation Repository*. 2926. https://ir.lib.uwo.ca/etd/2926

Orpinas, P., and Frankowski, R. (2001). The Aggression Scale: A Self-Report Measure of Aggressive Behavior for Young Adolescents. *Journal of Early Adolescence, 21*, 50-67.

Paulhus, D. L. (2014). Toward a taxonomy of dark personalities. *Current Directions in Psychological Science*, *23*(6), 421-426. doi:10.1177/0963721414547737

Pincus, A. L., Ansell, E. B., Pimentel, C. A., Cain, N. M., Wright, A. G., & Levy, K. N. (2009). Initial construction and validation of the Pathological Narcissism Inventory. *Psychological Assessment*, *21*(3), 365-379. doi:10.1037/a0016530

Pornari, C. D., & Wood, J. (2010). Peer and cyber aggression in secondary school students: the role of moral disengagement, hostile attribution bias, and outcome expectancies. *Aggressive Behavior*, *36*(2), 81-94. doi:10.1002/ab.20336

Pozzoli, T., Gini, G., & Thornberg, R. (2016). Bullying and defending behavior: The role of explicit and implicit moral cognition. *Journal of School Psychology*, *59*, 67-81. doi:10.1016/j.jsp.2016.09.005

Rawhide. (2017, October 31). Teen cyberbullying and social media use on the rise. Retrieved from http://www.rawhide.org/blog/wellness/teen-cyberbullying-and-social-media-use-on-the-rise/

Repacholi, B., Slaughter, V., Pritchard, M., & Gibbs, V. (2003). *Theory of Mind, Machiavellianism, and social functioning in childhood*. Retrieved from

https://www.researchgate.net/profile/Virginia_Slaughter/publication/37627377_Theory_of_mind_Machiavellianism_and_social_functioning_in_childhood/links/5408f7240cf2187a6a6ea666/Theory-of-mind-Machiavellianism-and-social-functioning-in-childhood.pdf

Ronningstam, E. (2017). Intersect between self-esteem and emotion regulation in narcissistic personality disorder - implications for alliance building and treatment. *Borderline Personality Disorder and Emotion Dysregulation*, *4*(1). doi:10.1186/s40479-017-0054-8

Runions, K. C., & Bak, M. (2015). Online moral disengagement, cyberbullying, and cyber-Aggression. *Cyberpsychology, Behavior, and Social Networking*, *18*(7), 400-405. doi:10.1089/cyber.2014.0670

Schwartz, S. J., Beyers, W., Luyckx, K., Soenens, B., Zamboanga, B. L., Forthun, L. F., … Waterman, A. S. (2010). Examining the light and dark sides of emerging adults' identity: A study of identity status differences in positive and negative psychosocial functioning. *Journal of Youth and Adolescence*, *40*(7), 839-859. doi:10.1007/s10964-010-9606-6

Schwartz, S. J., Klimstra, T. A., Luyckx, K., Hale, W. W., Frijns, T., Oosterwegel, A., … Meeus, W. H. (2011). Daily dynamics of personal identity and self-concept clarity. *European Journal of Personality*, *25*(5), 373-385. doi:10.1002/per.798

Seidman, G. (2014). Expressing the "True Self" on Facebook. *Computers in Human Behavior*, *31*, 367-372. doi:10.1016/j.chb.2013.10.052

Sest, N., & March, E. (2017). Constructing the cyber-troll: Psychopathy, sadism, and empathy. *Personality and Individual Differences*, *119*, 69-72. doi:10.1016/j.paid.2017.06.038

Suler, J. (2004). The online disinhibition effect. *Cyber Psychology & Behavior*, *7*(3), 321-326. doi:10.1089/1094931041291295

Tidwell, L. C., & Walther, J. B. (2002). Computer-mediated communication effects on disclosure, impressions, and interpersonal evaluations: Getting to know one another a bit at a time. *Human Communication Research*, *28*(3), 317-348. doi:10.1111/j.1468-2958.2002.tb00811.x

Tisak, M. S., Tisak, J., Goldstein, S. E. (2006). Aggression, delinquency, and morality: A social-cognitive perspective. M. Killen, J. Smetana (Eds.), *Handbook of moral development,* (pp. 611-632). Mahwah, NJ: Lawrence Erlbaum Associates.

Vaknin, S. (2015). *Malignant self love: Narcissism revisited* [Paper]. Retrieved from http://samvak.tripod.com/thebook.html

Van den Bos, W., & Hertwig, R. (2017). Adolescents display distinctive tolerance to ambiguity and to uncertainty during risky decision making. *Scientific Reports*, *7*. doi:10.1038/srep40962

Van Geel, M., Toprak, F., Goemans, A., Zwaanswijk, W., & Vedder, P. (2017). Are youth psychopathic traits related to bullying? Meta-analyses on callous-unemotional traits, narcissism, and impulsivity. *Child Psychiatry & Human Development*, *48*(5), 768-777. doi:10.1007/s10578-016-0701-0

Viding, E., Blair, R. J., Moffitt, T. E., & Plomin, R. (2005). Evidence for substantial genetic risk for psychopathy in 7-year-olds. *Journal of Child Psychology and Psychiatry*, *46*(6), 592-597. doi:10.1111/j.1469-7610.2004.00393.x

Weir, K. F., & Jose, P. E. (2010). The perception of false self scale for adolescents: reliability, validity, and longitudinal relationships with depressive and anxious symptoms. *British Journal of Developmental Psychology*, *28*, 393-411. doi:10.1037/t56207-000

Wright, M. F. (2014). Predictors of anonymous cyber aggression: The role of adolescents' beliefs about anonymity, aggression, and the permanency of digital content. *Cyberpsychology, Behavior, and Social Networking*, *17*(7), 431-438. doi:10.1089/cyber.2013.0457

Appendix A: Scale Items and Instructions

Online Disinhibition Scale Items

Items were administered with the following instructions: 'How much do you agree with the following statements.' All responses were collected on a 5-point Likert scale with anchors: 1 = *Strongly Disagree* to 5 = *Strongly Agree*

Online Disinhibition
- I feel safer expressing negative thoughts and feelings online than in person.
- The anonymity of the online environment influences the way I express myself online.
- I say/write/post comments online that I would not say in person.
- I am aware when I am being hurtful online, but do it anyway.
- I often feel guilty after posting something negative online.

Cyber Aggression Scale Items

Items were administered with the following instructions: 'In the last month (4 weeks/30 days) how often have you engaged in the following cyber behaviors:'

All responses were collected on a 5-point Likert scale with anchors: 1 = *Never* to 5 = *7 or more times*

Cyber Aggression
- Posted something on social media to anger or make fun of someone.
- Made mean or negative comments on someone's photos, updates, or tags to make that person feel bad or for others to join in and/or laugh.
- Spread rumors about someone online.
- Edited a photo or created a meme making fun of something and then posted it online for others to see.

- Sent someone an instant message to make him or her angry or to make fun of them.
- Taken an embarrassing picture of someone and posted it online without their permission.
- Posted something online about someone else to make others laugh.

Manuscript Highlights

- Adolescent narcissism, sadism, psychopathy, and false self perceptions predict online disinhibition.

- Perceptions of false self were found to be a significant predictor of cyber aggression when mediated by online disinhibition.

- Sadistic traits and online disinhibition were found to be significant direct predictors of cyber aggression.

- All three dark traits are predictive of cyber aggression through the indirect role of increased disinhibition.